INSTRUCTIONS FOR THE COMPARATIVE CASE IDENTIFICATION TOOL

TIMOTHY PRESCOTT AND BRIAN URLACHER

Brief Instructions

- Open `index.html`, click "Choose file", and upload a `.csv` file to analyze.
- The `.csv` file should have case names in the first column and variable names in the first row.
- Any non-numeric data is assumed to be categorical, and will be converted to natural numbers (0, 1, 2, ...) in the order encountered.
- Deteremine which variables should be ignored, controlled for, declared independent, and declared dependent.
- *Plot* will display the Principle Component for each axis.
- *Table* will compute the distances between the various cases, and convert those distances into a score.
- *Download results* will download a case by case `.csv` calculation of the three distances and two design scores.

Further Information

- All distances are Mahalanobis distances using the Moore-Penrose pseudo-inverse of the covariance matrix. This means a variable with a large standard deviation contributes less to the distance calculation than a variable with a small standard deviation. It also means that highly correlated variables contribute less to the distance. Using the pseudo-inverse allows the data to be linearly dependent.
- In a Most Similar design, dependent and independent variables should be far apart while control variables should be close together. The Similarity score of two cases is

$$\frac{\sqrt{(\text{Dependent distance})}\sqrt{(\text{Independent Distance})}}{\text{Control Distance}}.$$

Date: March 2, 2015.

- In a Most Different design, dependent and independent variables should be close while control variables should be far apart. The Difference score of two cases is
$$\frac{\text{Control Distance}}{(\text{Dependent distance}) + (\text{Independent Distance})}.$$
- Because categorical information is converted in the order encountered, more than two categories will cause CCIT to assume a linear relationship: the second category will come between the first and third categories. You should consider converting the categories to numerical values to make that relationship explicit.
- If you have a large number of cases (more than one thousand or so), you may want to use the slower method. The downside is that the table won't scroll smoothly and downloading won't be allowed. The upside is that these activities would have crashed the app.

Browser Specific Shortcomings

- Firefox and Internet Explorer won't work with the big data option (to clarify, the big data option works, but if the data is big enough to require it, the browser will crash).
- Safari on a Mac won't download the file. Safari will instead open the file in a new tab, and you must then save the data.
- Apple stopped developing Safari on Windows before they implemented file uploading. Consequently, you can't upload a `.csv`, and the app won't work.

Acknowledgements

This app makes use of the libraries jQuery, jQuery UI, jQuery SVG, Papa Parse, Numeric Javascript, and FileSaver.js.

Timothy Prescott and Brian Urlacher
E-mail address: timothy.prescott.und@gmail.com

E-mail address: brian.urlacher@business.und.edu

University of North Dakota